Bovine Enterovirus Type 2. Complete Genomic Sequence and Molecular Modeling of a Reference Strain and a Wild Type Isolate from Endemically Infected US Cattle.

S. D. Goens, S. Botero, A. Zemla, C. E. Zhou, M. L. Perdue

March 1, 2004

Bovine enterovirus type 2. Complete genomic sequence and molecular modeling of the reference strain and a wild type isolate from endemically infected US cattle.

Text and figure submissions to collaborators:

In Methods section:

Modeling of Capsid Proteins

Capsid proteins from BEV-2 strains PS-87 and Wye-3A were modeled using the AS2TS homology modeling system [1], by which protein sequences are compared against all PDB entries, and those with the highest homology are evaluated as suitable templates. PDB entry 1BEV had the highest sequence identity for each of the BEV-2 capsid proteins. In regions where residues had been omitted from the x-ray crystallography data, additional analyses using Local-Global Alignment (LGA) software [2] and secondary structure prediction using Psipred [3] were done in order to attempt completion of the templates. Models for the capsid proteins of strains PS-87 and Wye-3A were then constructed using the modified templates. Again using 1BEV as a template, protomers for PS-87 and Wye-3A were assembled and models were checked for possible clashes between residues on adjacent chains.

In Results section:

Modification of templates for each 1BEV chain.

	1BEV VP-1. Because 13 N-terminal residues of 1BEV_1 (VP-1) were not provided in the PDB entry, and because temperature factors for atoms of residues 14-17 were very high (above 60), we performed additional analyses in order to model this region. The results from Psipred secondary structure predictions indicate that N-terminal part of VP-1 can be considered as alpha helix. LGA structure comparison analyses of the closest homologues to 1BEV allowed us to identify a number of PDB entries of VP-1 proteins with alpha helical

N-terminus, with good sequence identity to 1BEV (above 35%), and the structures solved with relatively high resolution (2.90 Angstroms or less). For example, the atom coordinates of 1AYM_1 (VP-1 from human rhinovirus 16) were determined with the resolution of 2.15 Angstroms. We used LGA to superimpose the N-terminal regions of 1BEV_1 and 1AYM_1 and the results shown on Fig. 1 demonstrate good overall structural similarity (regions in green) between these two proteins, and possible helical conformation in first 17 N-terminal 1BEV_1 residues. Using sequence-based alignment (Fig. 2) as a guide and the N-terminus of 1AYM_1 as a template, we then modeled residues 4-17 of 1BEV_1 as alpha helix. Because terminal regions tend to be flexible, we generated several conformations in order to choose an acceptable spatial orientation for this helix; we examined each conformation in the context of the protomer complex to determine which would be in highest agreement (minimize clashes) with the 3 other proteins (VP-2, VP-3, and VP-4) (data not shown). Our modified 1BEV template was sufficient to completely model the VP-1 proteins for strains PS-87 and Wye-3A.

1BEV VP-2. Residues 1-4 of 1BEV_2 (VP-2) were not provided in the PDB entry. We were unable to model these residues due to lack of a PDB homolog for which coordinates were provided for the corresponding residues. The high temperature factors (low confidence) for residues 5-11, however, prompted us to adjust the structural assignment of these residues. Using the same approach described above for 1BEV_1, we identified 1AR7_2 as the best template for modeling this region and adjusted 1BEV_2 coordinates accordingly. This new conformation also was found to avoid clashes with reconstructed C-terminal disordered region of 1BEV_4.

1BEV VP-3. There were no missing residues in the PDB entry for 1BEV_3 (VP-3). Therefore, the VP-3 proteins for strains PS-87 and Wye-3A were modeled completely with high confidence by homology to 1BEV_3.

1BEV VP-4. We decided that 1BEV_4 (VP-4) was a poor template for modeling VP-4 of strains PS-87 and Wye-3A due to the lack of coordinate data for 28 of 68 residues and due to very high temperature factors for most of the residues in the PDB entry. Of 25 VP-4 homologs selected from PDB, we determined that entry 1HXS_4 of poliovirus (Mahoney strain), with 57% sequence identity to 1BEV_4 and solved with 2.2 Angstroms resolution, was the most complete and would provide the best template for modeling most regions of VP-4 proteins of strains PS-87 and Wye-3A. Fig. 3 summarizes LGA structural superpositions of these homologs. Pairwise comparison of each homolog with 1HXS_4 reveals excellent structural agreement among almost all of these homologs in all regions except for a region spanning approximately 17 residues beginning 8 positions from the N-terminus, in which many homologs are missing coordinate data or are poorly aligned, and a region corresponding to residues 40 to 50 of 1HXS_4 for which good alignment is evident among first 10 of the homologs (Fig. 3). It is also important to notice that all first 16 homologs (1hxs4,

1eah4, 2plv4, 1vbd4, 1pvc4, 1vba4, 1vbe4, 1vbc4, 1vbb4, 1asj4, 1ar64, 1d4m4, 1h8tD, 1ev14, 1oopD, 1jew4) have very similar overall structural conformation (mostly green and yellow areas including terminal regions). In order to address structural ambiguity and to suitably modify template 1BEV_4, we targeted further analysis to the flexible loop region of residues 8-24. When examining the modified 1BEV_4 template (based on 1HXS_4) in the context of the protomer, we discovered a clash between the N-terminus of 1BEV_2 and the C-terminus of 1BEV_4 (Fig. 4), and repositioned N-terminal residues of 1BEV_2 accordingly (cyan residues in Fig. 4). It may be that electron density map data corresponding to what we believe are C-terminal residues 64-68 of 1BEV_4 (blue residues alongside gray in Fig. 4) were mistakenly assigned to N-terminal residues 5-9 of 1BEV_2 (blue residues in Fig. 4), so we adjusted template 1HXS_4 to reflect this structural re-assignment.

Modeling of BEV-2 capsid proteins for strains PS-87 and Wye-3A.

The modified templates were used to construct homology models for each capsid protein of strains PS-87 and Wye-3A. The side chains from identical residues were incorporated to the models automatically, and the remaining side chain atoms were calculated using the program SCWRL [4]. Loop regions: were constructed using locally developed LGA_modeler program. In VP-1 protein there are four "ins/del" regions where 1BEV template differs from our strains: 86-92 (R1), 127-138 (R2), 226-230 (R3), 266-273 (R4), and in VP-3 we have one "ins" region: 180-187 (R5). All these regions are located on the surface of 1BEV protomer. Fig. 5 shows the complete protomer for strain PS-87, and the corresponding R1-R5 regions are colored in red. Virus particles are composed of 60 copies each of the four VP1-VP4 structural proteins. As it is shown on the Fig. 6 only VP1, VP2, and VP3 are exposed at the viral surface.

Figures:

Fig. 1. Structure-based alignment (first 40 residues from N-terminal part only; see Fig.2.) of 1AYM_1 (VP-1) and 1BEV_1 (VP-1) using LGA. Backbone representation of 1AYM_1 (thick) and 1BEV_1 (thin) is shown. In cyan is the N-terminal alpha helix (residues 1-14) of 1AYM_1, and in red is the high temperature region (residues 14-17) of 1BEV_1. Green indicates the remaining superimposed N-terminal residues of both proteins (15-36 of 1AYM_1 and 18-40 of 1BEV_1).

```
Conf:      9767889999988651145028622465756675577787...
```

```
Psipred:  CCHHHHHHHHHHHHHHHCEEECCCCCCCCCCCCCCCCCCCCCC...
1bev_1:   NDPGKMLKDAIDKQAAGALVAGTSTSTHSVATDSTPALQA...
1aym_1:   ---NPVERYVDEVLNEVLVVPNINQSHPTTSN-AAPVLDA...
```

Fig. 2. Sequence-structure based alignment of N-terminal regions of 1AYM_1 and 1BEV_1, and Psipred secondary structure prediction of N-terminal of 1BEV_1 region. Red, green, and cyan correspond to same residues depicted in Fig. 1. Purple indicates residues missing in entry 1BEV_1, and in red are colored residues solved in 1BEV_1 with low resolution (hight temperature factor).
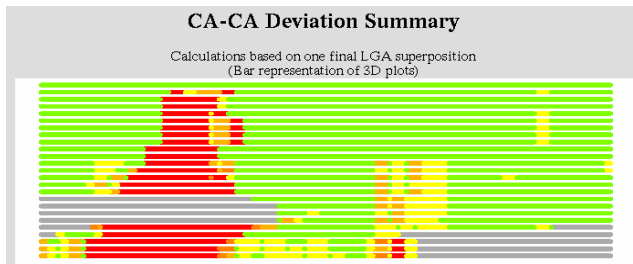

Fig. 3. Summary of structural alignments of 25 VP-4 PDB entries with entry 1HXS_4. Top line represents 1HXS_4. 1BEV_4 is fifth line from the bottom (see Fig. 4). Green indicates very good structural alignment (residues superimposed below 1.50 Angstroms); yellow indicates good alignment (below 3.00 Angstroms); orange indicates poor alignment (below 4.50 Angstroms), brown indicates very poor (below 6.00 Angstroms), and residues not aligned or missing in PDB are in red. Terminal residues that are completely missing in PDB entry are gray.
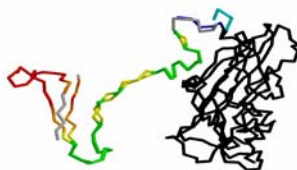

Fig. 4. Positioning of VP-2 and VP-4 in protomer context. 1BEV VP-2 (in black), 1HXS VP-4 superimposed with 1BEV VP-4 (coloring scheme the same as in fifth bar from the bottom in Fig. 3). Cyan indicates N-terminus of 1BEV_2 template modified (based on 1AR7_2) to avoid clash with C-terminus of VP-4. Original 1BEV VP-2 N-terminal region 5-9 as deposited in PDB is shown in blue (it overlaps with C-terminal gray end 64-68 from VP-4).
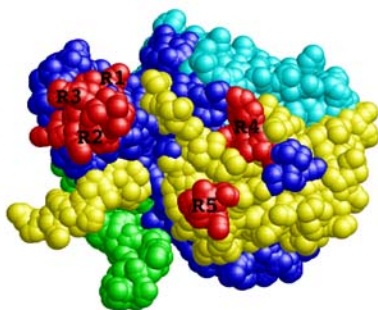
Fig. 5. 3D model of protomer for BEV-2 strain PS-87 based on modified 1BEV template. VP-1 (blue), VP-2 (cyan), VP-3 (yellow), VP-4 (green). Insertion and deletion regions R1-R5 are in red.
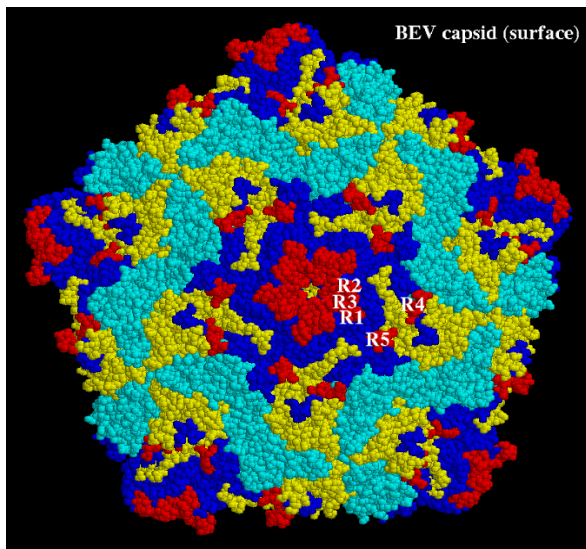

BEV capsid (surface)

Fig. 6. The surface of the BEV capsid (pentamer subunit) where protein VP1 is in blue, VP2 in cyan, and VP3 in yellow. Protein VP4 is not visible because it is completely buried under the surface. Insertion and deletion regions R1-R5 are colored in red. Regions R1-R3 and R4 are located on the rims of the observed canyon (see [5]) with R5 lying in its base.

References:

[1] - Zemla, A., Slezak, T., Barsky, D. and Sawicka, D.: "Automated 3D Protein Structure Predictions Based on Sensitive Identification of Sequence Homology", LLNL project: LDRD 02-LW-003, 2002.
[2] - Zemla, A.: "LGA - a method for finding 3D similarities in protein structures", Nucleic Acids Research, Vol. 31, No. 13, 2003, pp. 3370-3374.
[3] - Jones, D.T.: "Protein secondary structure prediction based on position-specific scoring matrices", J. Mol. Biol. 292, 1999, pp. 195-202.
[4] - Bower, M.J., Cohen, F.E., Dunbrack, R.L. Jr.: "Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool", J. Mol. Biol. 267, 1997, pp. 1268-1282.
[5] - Smyth, M., Tate, J., Hoey, E., Lyons, C., Martin, S., Stuart, D.: "Implications for viral uncoating from the structure of bovine enterovirus", Nat. Struct. Biol., 1995, 2(3), pp. 224-31.